## IN THE CLAIMS

Please amend the claims as follows:

1. (Withdrawn) A method for processing a memory instruction, the method comprising:
obtaining a memory instruction;
obtaining one or more memory address operands;
creating a virtual memory address using the one or more memory address operands;
translating the virtual memory address into a physical memory address; and
executing the memory instruction using a cache controller,
wherein the cache controller uses the memory instruction and the physical memory address to determine whether to access a portion of a local or a remote cache.

2. (Withdrawn) The method of claim 1, wherein the obtaining of the memory instruction includes obtaining a scalar memory instruction.

3. (Withdrawn) The method of claim 1, wherein the obtaining of the memory instruction includes obtaining a memory instruction selected from a group consisting of a scalar load instruction, a scalar store instruction, a prefetch instruction, a synchronization instruction, and an atomic memory operation (AMO) instruction.

4. (Withdrawn) The method of claim 1, wherein the obtaining of the memory instruction includes obtaining a scalar store instruction, and wherein the method further comprises obtaining scalar store data.

5. (Withdrawn) The method of claim 1, wherein the translating includes translating the virtual memory address into a physical memory address using a translation look-aside buffer (TLB).

6. (Withdrawn) The method of claim 1, wherein the method further comprises committing the memory instruction before execution.

7. (Previously Presented) A computerized method for accessing data in a memory system having a local cache and a higher level cache, comprising:

obtaining a memory request;

storing the memory request in an Initial Request Queue (IRQ); and

processing the memory request from the IRQ by a cache controller, wherein processing includes:

determining whether the memory request hits in the local cache;

determining whether a portion of an address associated with the memory request matches one or more partial addresses in a Force Order Queue (FOQ), wherein the FOQ stores a memory request that is pending to the higher level cache;

if the portion of an address associated with the memory request does not match the one or more partial addresses in the FOQ and, at the same time, the memory request hits in the local cache, servicing the memory request immediately using data in the local cache;

if the portion of an address associated with the memory request does not match the one or more partial addresses in the FOQ and, at the same time, the memory request misses in the local cache, adding the memory request to the FOQ, allocating a cache line in the local cache corresponding to the local cache miss and servicing the memory request using data received from the higher level cache; and

if the portion of an address associated with the memory request matches the one or more partial addresses in the FOQ, preventing the memory request from being satisfied in the local cache, wherein preventing includes adding the memory request to the FOQ and servicing the memory request using data received from the higher level cache.

8. (Original) The computerized method of claim 7, wherein the obtaining of the memory request includes obtaining a memory load or a memory store request.

9-13. (Canceled)

14. (Previously Presented) The computerized method of claim 7, wherein determining whether a portion of an address associated with the memory request matches one or more partial addresses in an Force Order Queue (FOQ) includes processing the memory request in the FOQ when local cache processing is bypassed.

15. (Previously Presented) The computerized method of claim 7, whercin determining whether a portion of an address associated with the memory request matches one or more partial addresses in an Force Order Queue (FOQ) includes processing the memory request in the FOQ when the memory request includes a synchronization request that causes local cache processing to be bypassed.

16. (Withdrawn) A computerized method, comprising:

adding a set of memory instructions of a first type to a memory instruction container;

executing the set of memory instructions of the first type from the memory instruction container;

receiving a memory instruction of a second type;

receiving a synchronization instruction that temporarily blocks further execution of memory instructions of thc first type from the memory instruction container;

adding an additional set of memory instructions of the first type to the memory instruction container;

executing the memory instruction of the second type; and

upon such execution of the memory instruction of the second type, executing the additional set of memory instructions of the first type from the memory instruction container.

17. (Withdrawn) The computerized method of claim 16, wherein the adding of the set of memory instructions of the first type to the memory instruction container includes adding the set of memory instructions of the first type to a memory instruction queue.

18. (Withdrawn) The computerized method of claim 16, wherein the adding of the set of memory instructions of the first type to the memory instruction container includes adding a set of scalar memory instructions to the memory instruction container.

19. (Withdrawn) The computerized method of claim 16, wherein the executing of the set of memory instructions of the first type from the memory instruction container includes accessing a local or remote cache unit.

20. (Withdrawn) The computerized method of claim 16, wherein the receiving of the memory instruction of the second type includes receiving a vector memory instruction.

21. (Withdrawn) The computerized method of claim 16, wherein the executing of the memory instruction of the second type includes accessing a remote cache unit, and updating a local cache unit with contents from the remote cache unit.

22. (Previously Presented) A scalar processor, comprising:

a local cache;

an Initial Request Queue (IRQ); and

a cache controller having a Force Order Queue (FOQ), wherein the FOQ stores a scalar memory request that missed in the local cache and is pending to a higher level cache;

wherein the IRQ buffers a scalar load/store memory request having a scalar load/store instruction and its one or more associated addresses and sends the scalar load/store memory request to the cache controller and the local cache;

wherein, when a portion of the one or more associated addresses of the scalar load/store memory request does not match one or more partial addresses in the FOQ and, at the same time, the scalar load/store memory request hits in the local cache, the local cache services the scalar load/store memory request received from the IRQ;

wherein, when the portion of the one or more associated addresses of the scalar load/store memory request does not match the one or more partial addressed in the FOQ and, at the same time, the scalar load/store memory request misses in the local cache, the scalar load/store

memory request is added to the FOQ, one or more lines in the local cache are allocated for cache line replacement, and the scalar load/store memory request is passed to the higher level cache; and

wherein, when the portion of the one or more associated addresses of the scalar load/store memory request matches the one or more partial addresses in the FOQ, the scalar load/store memory request is added to the FOQ and the scalar load/store memory request is passed to the higher level cache.

23. (Canceled)

24. (Previously Presented) The scalar processor of claim 22, wherein the scalar processor further includes a scalar load/store unit, wherein the scalar load/store unit includes an address generator to generate one or more physical addresses from the one or more associated addresses of the scalar load/store command.

25. (Original) The scalar processor of claim 24, wherein the address generator generates the one or more physical addresses using a translation look-aside buffer (TLB).

26. (Withdrawn) A processing unit, comprising:

a remote cache interface; and

a scalar processor having a local cache,

wherein the scalar processor dispatches a memory instruction, obtains one or more address operands, creates a virtual address from the one or more address operands, translates the virtual address into a physical address using a translation look-aside buffer (TLB), and executes the memory instruction using a cache controller, the cache controller determining whether to obtain data from the local cache or to allocate one or more lines of the local cache through the remote cache interface.

27. (Withdrawn) A processing unit, comprising:

a cache interface;

a vector processor; and

a scalar processor,

wherein the scalar processor processes a plurality of scalar memory instructions in a memory instruction container,

wherein the vector processor receives a vector memory instruction,

wherein the scalar processor receives a synchronization instruction and temporarily blocks further processing of scalar memory instructions in the memory instruction container,

wherein the vector processor processes the vector memory instruction by accessing a cache through the cache interface, and

wherein the scalar processor continues to process further scalar memory instructions in the memory instruction container after the vector processor has processed the vector memory instruction.

28. (Withdrawn) The processing unit of claim 27, wherein the memory instruction container comprises a memory instruction queue.

29. (Currently Amended) A scalar processor, comprising:

a local cache;

an Initial Request Queue (IRQ); and

a plurality of cache controllers, wherein each cache controller includes a Force Order Queue (FOQ), wherein ~~the~~ each FOQ receives scalar memory requests from the IRQ and stores [[a]] the scalar memory ~~request that~~ requests in the FOQ if the scalar memory requests missed in the local cache and ~~is~~ are pending to a higher level cache;

wherein the IRQ buffers a scalar load/store memory request having a scalar load/store instruction and its one or more associated addresses and sends the scalar load/store memory request to the local cache and to one of the plurality of cache controllers corresponding to the one or more associated addresses of the scalar load/store memory request;

wherein, when a portion of the one or more associated addresses of the scalar load/store memory request received by one of the FOQs does not match one or more partial addresses in the FOQ and, at the same time, the scalar load/store memory request hits in the local cache, the local cache services the scalar load/store memory request received from the IRQ;

wherein, when the portion of the one or more associated addresses of the scalar load/store memory request does not match the one or more partial addressed in the FOQ and, at the same time, the scalar load/store memory request misses in the local cache, the scalar load/store memory request is added to the FOQ, one or more lines in the local cache are allocated for cache line replacement, and the scalar load/store memory request is passed to the higher level cache; and

wherein, when the portion of the one or more associated addresses of the scalar load/store memory request matches the one or more partial addresses in the FOQ, the scalar load/store command is added to the FOQ and the scalar load/store memory request is passed to the higher level cache.

30. (Previously Presented) The scalar processor of claim 29, wherein the FOQ includes a FOQ index array, wherein the FOQ index array contains a copy of indices and control information for the FOQ entries.

31. (Previously Presented) The scalar processor of claim 29, wherein the FOQ is divided logically into a first and second queue, wherein the first queue monitors scalar memory requests to the higher level cache and the second queue monitors scalar memory requests that are serviced by the higher level cache but not yet written to the local cache.

32. (Previously Presented) The scalar processor of claim 22, wherein the FOQ is divided logically into a first and second queue, wherein the first queue monitors scalar memory requests to the higher level cache and the second queue monitors scalar memory requests that are serviced by the higher level cache but not yet written to the local cache.